Labor Demand

Daniel S. Hamermesh

PRINCETON UNIVERSITY PRESS
PRINCETON, NEW JERSEY

Published by Princeton University Press, 41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press,
Chichester, West Sussex

Library of Congress Cataloging-in-Publication Data

Hamermesh, Daniel S.
Labor demand / Daniel S. Hamermesh.
p. cm.
Includes bibliographical references and index.
1. Labor demand. 2. Labor market. I. Title
HD5706.H36 1993 331.12'3—dc20 92-23775

ISBN 0-691-04254-3
ISBN 0-691-02587-8 (pbk.)

This book has been composed in Linotron Palatino

Princeton University Press books are printed on acid-free paper and meet the guidelines for permanence and durability of the Committee on Production Guidelines for Book Longevity of the Council on Library Resources

First Princeton Paperback printing, 1996

Printed in the United States of America

10 9 8 7 6 5 4 3

To my parents, with love and respect

CHAPTER TWO

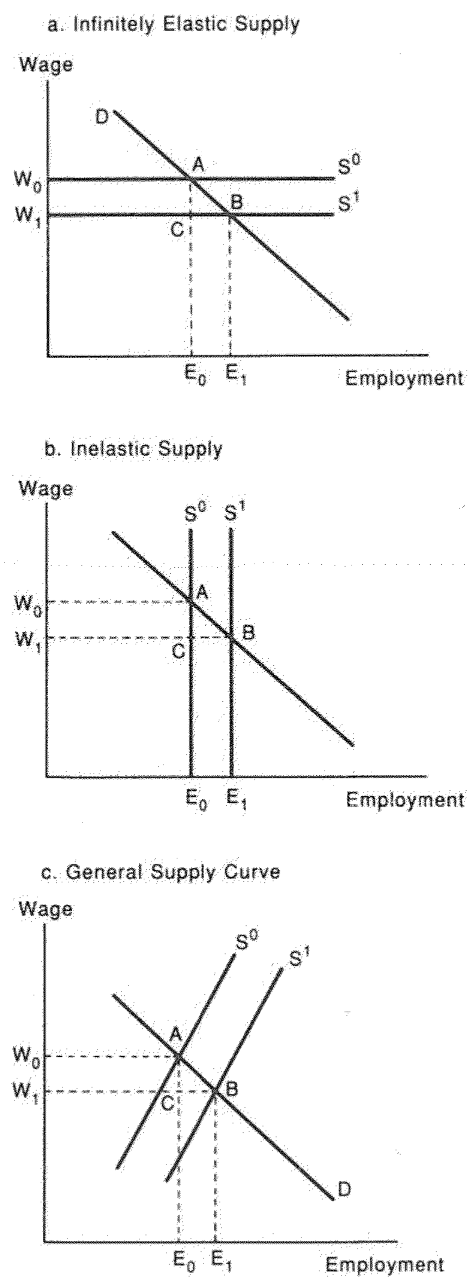
The Static Theory of Labor Demand

I. Introduction

In this chapter I demonstrate how parameters describing employers' long-run demand for labor can be inferred from data characterizing their employment, wages, product demand, and in some cases the prices and quantities of other inputs. Much of the exposition in Sections II–V is the standard neoclassical theory of factor demand, in which the effects on factor demand of small changes are analyzed. The purpose is not, however, to rehash this theory, but rather to show that it can be used to infer parameters of interest. Toward that end I spend a substantial amount of time indicating how the theory can be specified explicitly to enable one to infer the structure of production. More mathematical complexity can be found in Varian (1984); still more is available in the essays in Fuss and McFadden (1978).

The entire discussion assumes that a demand curve exists at the level of the firm. There is a longstanding controversy over the existence of an aggregate production function, and by inference therefore an aggregate labor-demand curve; but there is no long history of objections to the notion of a firm-level labor-demand relationship (Harcourt 1972). There are more recent objections, not so much to the underlying theoretical notion but rather to the usefulness of the construct in describing employment-wage outcomes (e.g., Oswald 1985). I discuss some of these objections in detail in Chapter 9 in the context of applying the results to aggregate labor markets. In the end, though, like any other internally consistent theory, its validity rests on its usefulness in describing measurable real-world phenomena.

For considering the appropriate form of the theory to use in deriving estimating relationships, there are two essentially polar ways of viewing labor demand. Neither is always correct; neither may ever be entirely correct. But both are useful for bracketing the likely responses of wages and employment to exogenous shocks. The first takes the view from the level of the individual employer that the wage is, in most cases, exogenous. Consider the firm shown in Figure 2.1a. It views supply as infinitely elastic at S^0, at a wage W_0. An increase in supply to S^1 produces a rise in employment from E_0 to E_1



2.1 Alternative Labor Supply Assumptions

because the wage has dropped to W_1. The entire direct effect of the shock is on wages, and that in turn produces an impact on employment that can be inferred if we know the slope of the labor demand curve, *AC/CB*.

Even in broader instances than at the level of the small firm—for example, in a unionized firm that operates on its demand curve, or where the supply of labor to a subsector is perfectly elastic—the wage can be viewed as unaffected by labor demand. In such cases being able to infer the magnitudes of wage elasticities of labor demand allows one to infer the effects of exogenous changes in wage rates on employers' labor demand. The impact of changes in the price of one type of labor on its employment and on the employment of other types of labor (cross-price effects) can be discovered using estimates of labor-demand relations alone. The assumption that wages (and other factor prices) are exogenous clearly presents problems when one tries to move from a firm, a small industry, or a unionized sector to the entire economy. If the particular type of labor under study has its wages *effectively* fixed by government, perhaps because it is paid an effective minimum wage; or if there is sufficient unemployment among workers of this type that the supply *to the market* is perfectly elastic, then it is reasonable to ask what will happen to employment when the exogenous wage is changed. Under this polar assumption about labor supply the static theory can be used to analyze the impact on employment of imposed changes in the wage of any one type (or types) of labor, in other factor prices, or in product demand.

In the alternative polar case one can in many instances assume that the employment of workers of a particular type is fixed (and determined solely by the completely inelastic supply of such workers to the market). Perhaps the economy is at full employment, so the supply curve of this (or any other) type of labor is completely inelastic, as shown by S^0 in Figure 2.1b. In this case employment is E_0, entirely determined by the supply of labor. The demand for labor determines the wage rate W_0 that workers of this type are paid. If, for example, there is an exogenous increase in the supply of labor, perhaps because of an increase in population in the group, or a greater taste for market work, the supply shifts out to S^1. The wage falls to W_1. Here again, knowing the slope *AC/CB* of the demand curve *D* provides the information needed to infer the effect of the shock on the market, but in this case the entire effect is on the wage rate. If one believes that, unless governments interfere by setting wage floors, labor markets must be characterized by full employment in the long run, this is the approach to use to analyze the comparative statics of supply shocks. Even with a less strictly new classical view of macroeconomics, this case is still clearly suited to analyzing the effects of supply shocks on those labor markets, and at those times, where and when there is full employment.

In general, neither perfectly elastic nor completely inelastic supply characterizes labor markets. Instead, the situation is such that supply has a positive finite slope, as shown by S^0 in Figure 2.1c. In this case an increase in supply to S^1 produces both an increase in employment and a reduction in the wage rate. Without knowing the slopes of *both* the supply and demand curves, one cannot infer the size of the changes in wages and employment. For a given supply shock, one can, though, still place an upper bound on the effect of the supply shock on employment (wages) using the demand curve alone, as in Figure 2.1a (2.1b).

How serious is ignoring supply by restricting the assumptions about supply to the two polar cases? There is no theoretical difficulty; the theory derived in this chapter applies regardless of what supply responses look like. The problem is the standard one of identification in econometrics, coupled with a desire to link the theory and estimation as closely as possible (Klein 1974, 137–45).

In many important instances problems of identification can be ignored. First, there are numerous cases where the first polar approach is appropriate because the wage of the particular type of labor is set by fiat. Second, that approach can also be useful in analyzing the effects of subsidies to (or taxes on) employment in particular occupations, for both economic theory and empirical work suggest that the elasticities of supply to particular occupations are quite high. Third, there is substantial evidence that for many groups of workers labor supply to the market is very inelastic (e.g., Killingsworth 1983). That being the case, the effect of population changes on those workers' wages can be studied using the second polar approach. For these reasons the theory can be viewed not only as specifying one of the two joint determinants of employment and wages. It is also directly useful in determining long-run impacts on employment or wages because in many cases the other is fixed.

Throughout Sections II–V I assume that labor is homogeneous in terms of hours worked and effort per hour. There is no consideration of the possible effects on the demand for labor of different types of workers or differences in their supply of hours, their willingness to expend effort or how these combine with the size of the work group to affect output. In Sections VI and VII I relax these assumptions. This departure necessitates a lengthy discussion of the nature of labor costs, for what makes the distinction between workers and hours

in production interesting is the effect of changes in the relative importance of the components of labor cost.

The purposes of this entire chapter are to exposit the theory of labor demand generally, and to show specifically how we can infer the effects of exogenous labor-market changes on the employment and/or wage rates of a group or groups of workers. The theoretical discussion of the static demand for labor is in Sections II–IV: demand for labor in the one-factor case, the two-factor case, and in the multifactor case. In the latter two sections I first derive the results generally, then proceed to specific functional forms. Section V examines the results in the context of nonprofit organizations. The theoretical discussion of the demand for workers and hours in Sections VI and VII is necessarily somewhat less formal. Partly the difference in treatment stems from the formal similarity between many of the issues treated there and those handled in Sections II–IV, and from the relative sparseness of the literature and narrowness of the topic. Partly, too, much of the literature has developed very specific models that provide few general insights about the issues of concern.

The focus throughout is on the relations between exogenous wage changes and the determination of employment, and between exogenous changes in inelastically supplied labor and the structure of relative wages. I generally assume that the typical firm maximizes profits, though Section V does analyze how the theory of labor demand is affected by alternative assumptions about what the enterprise maximizes. I also assume throughout that employers are perfect competitors in the labor market. Most of the discussion assumes that the employer is also a perfect competitor in the product market. While this latter assumption may not be correct, the analysis applies mutatis mutandis to employers who have some product-market power.

II. Labor Demand with One Input

Though the basic theorems of labor demand require assuming that there are at least two inputs into production, some very useful results can be derived when only one input is assumed. Included among these is a motivation for the downward-sloping labor-demand curves in Figures 2.1a–c.

Let L be the homogeneous labor input, W the nominal wage, and P the product price. In this section I assume that output is produced by a production function that transforms labor services into output, $\phi(L)$, with $\phi' > 0$, $\phi'' < 0$. In other words, there are diminishing returns to the single input, labor. This can be assumed to be a short-run production function in which all other input amounts are held constant. Assume for the moment that the firm is competitive in all markets. It attempts to maximize profits

$$\pi = P\phi(L) - WL,$$

which it does by setting

$$\phi'(L^*) - w = 0, \tag{2.1}$$

where $w = W/P$ is the real wage and L^* is the profit-maximizing demand for labor. Condition (2.1) is the standard rule that the profit-maximizing firm sets the value of the marginal product equal to the real wage. It yields a maximum, for $\phi'' < 0$. The result shows that for a firm that is competitive in the product market we need only consider changes in real factor prices.

The condition also leads us to infer the downward-sloping demand curve of the figures. Differentiating in (2.1) and rearranging terms:

$$\frac{dL^*}{dw} = \frac{1}{\phi''(L^*)} < 0. \tag{2.2}$$

The negative slopes of the demand curves in the figures are based on the concavity of the one-factor production function. The more rapidly diminishing are the returns to labor (the more negative is ϕ''), the steeper is the demand curve for labor.

If the product market is not perfectly competitive, profits become $P(\phi(L)) \cdot \phi(L) - WL$, with P now a decreasing function of output. The profit-maximizing demand for labor is now determined by $P'(L^*)\phi'(L^*)\phi(L^*) + P\phi'(L^*) - W = 0$, which, by multiplying the first term by P/P and remembering the definition of an elasticity, is

$$\phi'(L^*)[1 - \frac{1}{\eta}] = \frac{W}{P}, \tag{2.3}$$

where $\eta \geq 0$ is the absolute value of the elasticity of product demand. Notice that the only difference between (2.3) and (2.1) is that the inverse of the product demand elasticity is subtracted. The condition now states that the firm chooses employment by setting the marginal revenue product equal to the real wage. For a perfectly competitive firm, $\eta \to \infty$, so (2.3) reduces to (2.1). This derivation shows that, other things equal, labor demand is also more steeply sloped the less elastic is the demand for the product.

III. Labor Demand with Two Inputs

The important results that the labor-demand curve slopes downward and that the elasticity of product demand affects labor demand are not a useful theoretical basis for serious empirical research in this area. First, the assumption of only one input is patently unrealistic and leaves unanswered the question of why there should be diminishing returns to the (single) factor. Second, the crucial notion of factor substitution, which underlies most empirical work, is impossible to discuss when only one input is assumed.

Many useful insights beyond those of the previous section come from examining the demand for homogeneous labor when there is only one cooperating factor. The convention is to assume that capital services are the other factor, which makes sense given the role of those services as the second biggest component of value added in most industries. Many of the specific mathematical forms for the production and cost functions from which labor-demand functions are derived were initially developed for the two-factor case and make more economic sense applied to only two factors than generalized to several.

Assume that production exhibits constant returns to scale, as described by the linear homogeneous function F, such that:

$$Y = F(L,K),\ F_i > 0,\ F_{ii} < 0,\ F_{ij} > 0, \tag{2.4}$$

where Y is output, and K is homogeneous capital services. In this initial part of the derivation, I assume the firm maximizes profits

$$\pi = F(L,K) - wL - rK, \tag{2.5a}$$

where r is the exogenous price of capital services, and I assume the competitive product price is one. Maximizing (2.5a) yields

$$F_L = w \tag{2.5b}$$

and

$$F_K = r. \tag{2.5c}$$

The competitive firm sets the value of the marginal product of each factor equal to its price. The ratio of (2.5b) to (2.5c),

$$\frac{F_L}{F_K} = \frac{w}{r}, \tag{2.5d}$$

is the familiar statement that the ratio of the values of marginal products, the marginal rate of technical substitution, equals the factor-price ratio.

Allen (1938, 341) defines the elasticity of substitution between the services of capital and labor as the effect of a change in relative factor prices on relative inputs of the two factors, *holding output constant*. (Alternatively, it is the effect of a change in the marginal rate of technical substitution on the ratio of factor inputs, defined as an elasticity.) Intuitively, this elasticity measures the ease of substituting one input for the other when the firm can only respond to a change in one or both of the input prices by changing the relative use of the two factors without changing output. In the two-factor linear homogeneous case the elasticity of substitution is

$$\sigma = \frac{dln(K/L)}{dln(w/r)} = \frac{dln(K/L)}{dln(F_L/F_K)} = \frac{F_L F_K}{Y F_{LK}} \tag{2.6}$$

(Allen 1938, 342–43). By this definition σ is always nonnegative.

Following Allen (1938, 372–73), the price elasticity of labor demand with output and r constant is

$$\eta_{LL} = -[1-s]\sigma < 0, \tag{2.7a}$$

where $s = wL/Y$, the share of labor in total revenue. η_{LL} measures the *constant-output labor-demand elasticity*. Intuitively, η_{LL} is smaller (less negative) for a given technology σ when labor's share is greater, because there is relatively less capital toward which to substitute when the wage rises. Equation (2.7a) reflects the first of Marshall's four laws of derived demand, that the own-price elasticity is higher the more easily the other factor is substituted for labor.

The *cross-elasticity of demand for labor* in response to a change in the price of capital services is

$$\eta_{LK} = [1-s]\sigma > 0. \tag{2.7b}$$

The intuition for including [1-s] here is that if capital's share is very small, a 1 percent change in its price cannot induce a large percentage change in labor demand, because the possible change in spending on capital services is small relative to the amount of labor being used. Both (2.7a) and (2.7b) reflect substitution between inputs, the crucial element missing in the previous section.

When the wage rate increases, the cost of producing a given output rises. In a competitive product market a 1 percent rise in a factor price raises cost, and eventually product price, by that factor's share. This reduces the quantity of output sold. The *scale effect* is thus the factor's share times the product-demand elasticity. To obtain the total demand elasticities for labor, scale effects must be added to (2.7a) and (2.7b):

$$\eta'_{LL} = -[1-s]\sigma - s\eta, \tag{2.7a'}$$

and

$$\eta'_{LK} = [1-s][\sigma - \eta]. \tag{2.7b'}$$

The term $s\eta$ in (2.7a') reflects Marshall's second law of derived demand: Input demand is less elastic when the demand for the product is less elastic, as we saw in the one-factor case. Equation (2.7a') is the fundamental law of factor demand. It divides the labor-demand elasticity into substitution and scale effects.[1] It can be derived using the production-function analysis employed thus far; but the derivation is much simpler using cost functions, so that I delay it until those have been introduced.

The representations (2.7a') and (2.7b') are best thought of as describing effects on labor demand in competitive firms that have the same production function and demand elasticity, η, for the industry's product. These results and (2.7a) and (2.7b) are the most important in the theory of labor demand. (Clearly, if we are dealing with factor demand by one competitive firm that *alone* experiences a change in a factor price, η'_{LL} and η'_{LK} approach $-\infty$, since the drop [rise] in the factor price leads the firm to expand [contract] forever.)

Both (2.7a) and (2.7b), and (2.7a') and (2.7b'), are useful, depending on the assumptions one wishes to make about the problem under study. For competitive firms in a particular industry, which can expand or contract as the wage changes, scale effects on employment demand are relevant. In that case (2.7a') and (2.7b') are more appropriate for inferring the potential effects of changes in input prices. If the typical firm's output supply is constrained, or, more interestingly, if we wish to apply these definitions to an entire closed economy operating at full employment, (2.7a) and (2.7b) are the correct measures of the long-run effect on labor demand of changes in the wage rate and the price of capital services.

All of these measures assume that both labor and capital services are supplied elastically to the firm. If they are not, the increases in employment when the wage decreases cannot be complete: The labor that is demanded may not be available; and the additional capital services whose presence raises the marginal product of labor ($F_{LK} > 0$) also may not be. In such cases the demand elasticities are reduced (Hicks 1932, Appendix). The example of a limit on the supply of capital services illustrates Marshall's third law of input demand.[2] These cases may be important, but I ignore them in the discussion. I do, though, deal with the polar case that assumes that employment is fixed but wages are flexible.

A dual approach is based on cost minimization. At the start total cost is assumed to be the sum of products of the profit-maximizing input demands and the factor prices. Total cost is linear homogeneous in input prices (doubling all nominal prices just doubles total

[1] The discussion here is based on constant returns to scale. The case of input demand with decreasing returns to scale (obversely, increasing marginal cost) is discussed by Mosak (1938). Though the analysis clearly differs (and is more complex), the distinction between substitution and scale effects still applies.

[2] The fourth law, which is based on the conditions describing the other three, is that the demand for an input that accounts for a small share of costs will be less elastic, other things equal, because the scale effect will be very small. A good intuitive discussion of these laws is provided by Stigler (1987).

cost, regardless of the degree of homogeneity of the production function). It can be written as

$$C = C(w,r,Y),\ C_i > 0,\ C_{ij} > 0,\ i,j = w,r, \tag{2.8}$$

since the profit-maximizing input demands were themselves functions of input prices, the level of output, and technology. By Shephard's lemma (see Varian 1984, 54) the firm's demand for labor and capital can be recovered from the cost function (2.8) as

$$L^* = C_w, \tag{2.9a}$$

and

$$K^* = C_r. \tag{2.9b}$$

Taking the ratio of these two conditions,

$$\frac{L^*}{K^*} = \frac{C_w}{C_r}. \tag{2.9c}$$

Intuitively, the cost-minimizing firm uses inputs in ratios equal to their marginal effects on costs.

The forms (2.9) are particularly useful for estimation purposes, since they specify the inputs directly as functions of the factor prices and output. One can write (2.9a) as

$$L^* = L^d(w,r,Y), \tag{2.9a'}$$

which can be written in logarithmic form for easy estimation as a log-linear equation. In such a form it yields the constant-output elasticity of demand for labor, η_{LL}, the cross-elasticity of demand, η_{LK} and the employment-output elasticity. Similarly, many researchers have rewritten (2.9c) as

$$\frac{L^*}{K^*} = l^d(w,r,Y). \tag{2.9c'}$$

Unlike (2.9a'), estimating (2.9c') does not provide direct measures of the demand elasticities.

Using equations (2.9) and the result that $C(w,r,Y) = YC(w,r,1)$ if Y is linear homogeneous, the elasticity of substitution can be derived:

$$\sigma = \frac{CC_{wr}}{C_w C_r} \tag{2.10}$$

(see Uzawa 1962). The form one uses to measure σ, (2.6) or (2.10), should be dictated by convenience.

The constant-output *factor-demand elasticities* can be computed as

$$\eta_{LL} = -[1-m]\sigma, \tag{2.11a}$$

and

$$\eta_{LK} = [1-m]\sigma, \tag{2.11b}$$

where m is the share of labor in total costs. Since, by the assumptions characterizing perfect competition, factors are paid their marginal products, and since the production and cost functions are linear homogeneous, $m = s$, and (2.11a) and (2.11b) are equivalent to (2.7a) and (2.7b).

With this apparatus it is now easy to prove the fundamental law of factor demand, (2.7a'). Following Dixit (1976, 79), continue to assume constant returns to scale, so that we can treat the firm as an industry. Industry factor demands are just the right-hand sides of (2.9a) and (2.9b) multiplied by industry output. Under competition firms equate price, p, to marginal and average cost:

$$p = C.$$

Noting that if markets clear, so that output equals industry demand $D(p)$,

$$\frac{\partial L}{\partial w} = YC_{ww} + D'(p)C_w^2.$$

Because C is linear homogeneous, $C_{ww} = (-r/w)C_{wr}$. Substituting for C_{ww}, then from (2.10) for C_{wr}, and then for C_w and C_r from (2.9a) and (2.9b),

$$\frac{\partial L}{\partial w} = \frac{rK}{Y}\frac{\sigma L}{wC} + \frac{D'(p)L^2}{Y^2}.$$

To put this into the form of an elasticity, multiply both sides by pw/pL:

$$\eta'_{LL} = -\frac{rK}{pY}\sigma + \frac{pD'(p)}{Y}\frac{wL}{pY} = -[1-s]\sigma - s\eta,$$

by the definition of factor shares under linear homogeneity. This is (2.7a').

The production or cost functions can also be used to define some concepts that are helpful for studying markets where real factor prices are flexible and endogenous, but factor supplies are fixed (and because of the flexibility of input prices, the second polar case in the introduction to this chapter, are fully employed). The converse of asking, as we have, what happens to the single firm's choice of inputs in response to an exogenous shift in a factor price is to ask what happens to factor prices that the representative firm must pay in response to an exogenous change in factor supply, as in Figure 2.1.b. Define the *elasticity of complementarity* as the percentage responsiveness of relative factor prices to a 1 percent change in relative inputs:

$$c = \frac{\partial ln(w/r)}{\partial ln(K/L)}. \tag{2.12}$$

This is the inverse of the definition of σ. Thus,

$$c = \frac{1}{\sigma} = \frac{C_w C_r}{C C_{wr}} = \frac{Y F_{LK}}{F_L F_K}. \tag{2.13}$$

In this two-factor case with a linear homogeneous production technology, one can find the elasticities of substitution and of complementarity equally simply from the production and cost functions. Having found one, the other is immediately available.

With constant marginal costs, an assumption that is analogous to the assumption of constant output in (2.7a) and (2.7b), the *elasticities of factor price* (of the wage rate and the price of capital services) are defined as

$$\epsilon_{ww} = -[1-m]c, \tag{2.14a}$$

and

$$\epsilon_{rw} = [1-m]c. \tag{2.14b}$$

Equation (2.14a) states that the percentage decrease in the wage rate necessary to accommodate an increase in labor supply with no change in the marginal cost of the product is smaller when the share of labor in total costs is larger. This occurs because labor's contribution to costs—a decrease—must be fully offset by a rise in capital's contribution in order to meet the condition that marginal cost be held constant.

Consider now some examples of specific production and cost functions. These are the main specific forms that have been used to infer the sizes of the crucial parameters, σ, η_{LL}, and η'_{LL}, in empirical studies of various industries, labor markets, and economies.

A. Cobb-Douglas Technology

The production function is

$$Y = AL^{\alpha}K^{1-\alpha}, \tag{2.15}$$

where α is a parameter, and A is some scale parameter that I assume hereafter equals one. The marginal products are

$$\frac{\partial Y}{\partial L} = \alpha\frac{Y}{L}, \tag{2.16a}$$

and

$$\frac{\partial Y}{\partial K} = [1-\alpha]\frac{Y}{K}. \tag{2.16b}$$

Since the ratio of (2.16a) to (2.16b) is w/r if the firm is maximizing profits, taking logarithms and differentiating with respect to $ln(w/r)$ yields $\sigma = 1$. Equations (2.7a) and (2.7b) imply

$$\eta_{LL} = -[1-\alpha] \text{ and } \eta_{LK} = 1-\alpha.$$

Minimizing total costs subject to (2.15), one can derive the demand functions for L and K, and thus the cost function. The latter reduces to

$$C(w,r,Y) = Zw^{\alpha}r^{1-\alpha}Y, \tag{2.17}$$

where Z is a constant. Using Shephard's lemma for both L and K in this specific case, one can derive

$$\frac{L}{K} = \frac{\alpha}{1-\alpha}\frac{r}{w}. \tag{2.18}$$

Taking logarithms,

$$ln\left(\frac{L}{K}\right) = \alpha' + ln\left(\frac{r}{w}\right), \tag{2.18'}$$

where α' is a constant. This form is very easy to use for estimation. It is trivial to show in (2.18′) that $\sigma = 1$ and also that $c = 1$. Moreover, it is clear from (2.18) alone that $\eta'_{LL} = -1$.

While the Cobb-Douglas function is easily used, the severe restrictions on all the interesting parameters render it of little current interest, since the purpose usually is to discover the sizes of labor-demand elasticities, not to assume that they equal $-[1-\alpha]$ and -1. Its only real advantage, given current computing technology, is its simplicity in providing a theoretical basis for inferring the size of labor's contribution to output. Indeed, that was its original purpose (Douglas 1976).

B. Constant Elasticity of Substitution Technology

The linear homogeneous production function is

$$Y = [\alpha L^{\rho} + (1-\alpha)K^{\rho}]^{1/\rho}, \tag{2.19}$$

where α and ρ are parameters, $1 > \alpha > 0$, $1 \geq \rho \geq -\infty$. Marginal products are[3]

$$\frac{\partial Y}{\partial L} = \alpha\left(\frac{Y}{L}\right)^{1-\rho}, \tag{2.20a}$$

and

$$\frac{\partial Y}{\partial K} = [1-\alpha]\left(\frac{Y}{K}\right)^{1-\rho}. \tag{2.20b}$$

[3] The trick to derive (2.20a) and (2.20b) is to remember that, after having done the arithmetic, the numerator is just Y raised to the power $1-\rho$.

ED_000584A_00001563

Letting the ratio of (2.20a) to (2.20b) equal to the factor-price ratio, taking logarithms, and differentiating with respect to $ln(w/r)$ yields

$$-\frac{\partial ln(K/L)}{\partial ln(w/r)} = \sigma = \frac{1}{1-\rho}. \qquad (2.21)$$

The CES is sufficiently general that σ is free to fluctuate between 0 and ∞, so that one can infer its size and that of the η_{LL}.

Among special cases of the CES are: (1) the Cobb-Douglas function ($\rho = 0$, as is clear if one lets $\rho \to 0$ in (2.21)); (2) the linear function ($\rho = 1$). From (2.19) $F_{LK} = 0$ if $\rho = 1$, so that from its definition $\sigma \to \infty$. In this case L and K are perfect substitutes; and (3) the Leontief function ($\rho \to -\infty$), in which case output is the minimum function $Y = \min\{L, K\}$, and $\sigma = 0$, so the inputs are not substitutable at all.[4] The constant-output factor-demand elasticities follow immediately from the definitions and the recognition that α is labor's share of revenue if the factors are paid their marginal products.

The CES cost function can be derived (Ferguson 1969, 167) as

$$C = Y\left[\alpha^{\sigma}w^{1-\sigma} + [1-\alpha]^{\sigma}r^{1-\sigma}\right]^{1/(1-\sigma)},$$

where, as before, $\sigma = \dfrac{1}{[1-\rho]} \geq 0$. The demand for labor is

$$L = \frac{\partial C}{\partial w} = \alpha^{\sigma}w^{-\sigma}Y. \qquad (2.22)$$

Taking logarithms in (2.22) yields

$$ln\, L = \alpha'' - \sigma ln\, w + ln\, Y, \qquad (2.22')$$

where α'' is a constant. The form (2.22′) is very useful for estimation.

In these examples it is straightforward to derive c first, then to derive σ as its inverse. It is worth noting for later examples and for the multifactor case that c is more easily derived from equations (2.20) and the factor-price ratio (since w/r, the outcome, appears alone) than from (2.22) and the demand for capital. σ is more readily derived from the cost function, since the ratio L/K appears alone. Obviously, in the two-factor case the simple relation (2.13) allows one to obtain c or σ from the other; but the ease of initially obtaining c or σ differs depending on which function one starts with, a difference that is magnified in the multifactor case.

A variant on the CES function is the *variable elasticity of substitution function*, in which $\sigma = h(L/K)$, where h is some continuous function (e.g., Lovell 1973). This assumption maintains the linear homogeneity of the function in (2.19) while allowing the elasticity of substitution to change as the ratio of inputs changes. Probably because it is difficult to develop any intuition about h', and because it is not easy to use in estimating equations like (2.22′), this formulation has only rarely been used in studies of labor demand.

[4] The arithmetic that demonstrates this is in Varian (1984, 18).

Several other specific functional forms, the generalized Leontief form of Diewert (1971), the translog form (Christensen, Jorgenson, and Lau 1973), and the CES-translog of Pollak, Sickles, and Wales (1984), are second-order approximations to arbitrary cost or production functions. Like the variable elasticity function, each has the advantage over the CES function in the two-factor case that σ (or c) is not restricted to be constant, but instead depends on the values of the factor inputs or prices.

C. *Generalized Leontief*

This approximation specifies

$$C = Y\{a_{11}w + 2a_{12}w^{.5}r^{.5} + a_{22}r\}, \qquad (2.23)$$

where the a_{ij} are parameters. Applying Shephard's lemma to (2.23) for each input,

$$L/Y = a_{11} + a_{12}[w/r]^{-.5}, \qquad (2.24a)$$

and

$$K/Y = a_{22} + a_{12}[w/r]^{.5}. \qquad (2.24b)$$

As can be seen by taking the ratio of (2.24a) to (2.24b), in general σ depends on all three parameters and the ratio w/r. Equation (2.24a) is easily estimated in logarithmic form by itself or jointly with (2.24b), providing estimates of the constant-output labor-demand elasticity that vary with the ratio of input prices. If $a_{12} = 0$, (2.23) becomes a Leontief function (since the ratio of L to K is always a_{11}/a_{22}).

D. *Translog and CES-Translog*

The translog cost function is

$$ln\, C = ln\, Y + a_0 + a_1 ln\, w + [1-a_1]ln\, r + .5b_1[ln\, w]^2 + b_2[ln\, w][ln\, r] + .5b_3[ln\, r]^2, \qquad (2.25)$$

where a_1 and the b_i are parameters. Applying Shephard's lemma to the labor input, and taking the ratio of both sides to total costs,

$$s = a_1 + b_1 ln\, w + b_2 ln\, r. \qquad (2.26)$$

Here too σ depends on all parameters and both factor prices. If $b_i = 0$ for all i, the cost function reduces to a Cobb-Douglas technology.

Equation (2.26) alone is ideally suited for estimating purposes and provides all the available information about the structure of production (since $b_1 + b_2 = b_2 + b_3 = 0$, due to the linear homogeneity of the cost function in w and r).

The CES-translog is a variant of the translog function that replaces the terms a_1 and $[1 - a_1]$ with

$$ln\left\{a_1 w^{1-\sigma} + [1-a_1]r^{1-\sigma}\right\}^{1/[1-\sigma]}.$$

The equation for labor's share becomes

$$s = \frac{a_1 w^{1-\sigma}}{a_1 w^{1-\sigma} + [1-a_1]r^{1-\sigma}} + b_1 ln\ w + b_2 ln\ r. \qquad (2.26')$$

The only difference between this equation and (2.26) is in the first term, which specifies a nonlinearity that permits estimation of an additional parameter that allows somewhat more generality in the elasticities. This formulation takes off from the CES function, for if all the b_i are zero, the cost function reduces to that of a CES.

All three of these formulations may be useful for empirical work, even when written out as in (2.23) and (2.25). Each has the virtue of allowing flexibility and containing some simpler forms as special cases. That suggests that they should supplant the Cobb-Douglas and CES functions even for empirical work involving just two inputs.

Throughout this section we have assumed the production function is linear homogeneous. Linear homogeneous functions are a subset of *homothetic* functions. In this broader class factor demand is such that the ratio of inputs is independent of scale at each factor-price ratio. This assumption may not always make sense. For example, large firms may increase efficiency by using a more capital-intensive process than small firms at given w and r. Alternatively, a particular firm facing the same factor prices may combine resources more efficiently in different proportions as its scale of operations changes.[5]

In the general case heterotheticity means that the production function cannot be written as $Y = G(F[L,K])$, where G is monotonic and F is linear homogeneous. Still more restrictively, the cost function cannot be expressed as

$$C(w,r,Y) = C^1(Y) \cdot C^2(w,r).$$

If production is heterothetic, output is not separable from factor prices; instead, the effect of factor prices depends on the scale of output. Some special cases are useful for estimation; and heterothetic CES-type functions (Sato 1977) and translog forms (Berndt and Khaled 1979) have been used. The latter involves the addition of the terms

$$\delta ln\ Y \cdot ln\ w$$

and

$$[1 - \delta] ln\ Y \cdot ln\ r$$

to (2.25), which results in the addition of $\delta ln\ Y$ to (2.26). Alternatively, if one does not wish to impose a particular functional form, examining whether the term in output belongs in a loglinear version of (2.9c') provides a test for homotheticity. These additional generalizations of production are useful if one believes that demand parameters depend on scale *and if* the underlying data show sufficient variation to allow one to test for heterotheticity by, for example, including the terms in $ln(Y)$ in the estimating equations.

Throughout this and the next section the maintained assumption is that the product price is constant at 1. This also implies that w and r are measured in real units. One could just as easily replace F_L and F_K in (2.5b) and (2.5c) by PF_L and PF_K and derive the implications for labor demand of shifts in industry demand. These modifications produce only scale effects as long as F is homothetic. Indeed, generalizing still more by assuming that the firm is noncompetitive, so that the price the firm charges depends on its output, still produces only scale effects if F is homothetic. Thus unless one abandons homotheticity, the derivations of constant-output factor-demand and other elasticities in this chapter are generally applicable.

IV. Labor Demand with Several Inputs

The derivation of factor-demand relationships with more than two inputs is of general interest to economists and should be of particular interest to labor economists when labor is one of those inputs. In that case we can tell, for example, how employment or wages are affected when the price or quantity of any one of several other inputs changes. It is useful to labor economists when we disaggregate labor along some interesting dimension, for example, age, race, sex, education, immigrant status, skill, occupation. In that case the theory of production with several inputs allows us to infer how changes in the wage rate of one group of workers affect the demand for labor in other groups (following the first polar approach to studying demand, that factor prices are exogenous); or how changes in the supply of

[5] My favorite example of heterotheticity is leaf raking on my own campus. Three technologies are used: (1) one worker with a rake cleaning up leaves behind bushes and next to buildings; (2) one worker with a lawnmowerlike machine in small open areas; (3) one worker with a giant fan on a flatbed truck blowing leaves into huge piles whence they are vacuumed into a compactor mounted on another truck, used in large open areas.

one group affect th [redacted]
second approach, [redacted]
Mathematically [redacted]
generalization of t [redacted]
though, the gener [redacted]
lated aspect of fact [redacted]
is classified into o [redacted]
sidering a three-fa [redacted]
L_2, and L_3. One co [redacted]

$Y = F(G(L_1, L$ [redacted]

where F and G ar [redacted]
discussed above. [redacted]
L_1 and L_2 by the fu [redacted]
Far better to devis [redacted]
aggregation.

This problem, [redacted]
son why labor eco [redacted]
mand. For examp [redacted]
skilled workers o [redacted]
differ depending [redacted]
Nonseparability [redacted]
of labor subaggreg [redacted]
related the group [redacted]
amount of one ty [redacted]
the ease of substi [redacted]
trarily included i [redacted]
rect inferences ab [redacted]
factors (and about [redacted]

Consider a firm [redacted]
production functi [redacted]

$Y = f(X_1, . .$ [redacted]

Then the associa [redacted]
$. . . , X_N$, is

$C = g(w_1, . .$ [redacted]

where the w_i are [redacted]

$f_i - \lambda w_i = 0,$ [redacted]

and using the cos [redacted]

$X_i - \mu g_i =$ [redacted]

where λ and μ a [redacted]
The technologi [redacted]
librium conditio [redacted]

degree homogeneity of factor demands in all factor prices), at least one $\eta_{ij} > 0$, $j \neq i$. What makes the multifactor case interesting is that some of the η_{ij} may be negative for $j \neq i$. This means, for example, that an increase in the wage rate of one group of workers with output constant might reduce employment of one or more other groups of workers as well as that of the workers whose wage rate has increased.

The *partial elasticity of complementarity* between two factors is defined using the production function as

$$c_{ij} = \frac{Yf_{ij}}{f_i f_j} . \tag{2.35}$$

This definition is a straightforward generalization of (2.13). The c_{ij} show the percentage effect on w_i/w_j of a change in the input ratio X_i/X_j, holding marginal cost and other input quantities constant. They provide a general way of analyzing the effects implicit in the polar case illustrated by Figure 2.1b. Just as the σ_{ij} are not invariant to changes in relative factor prices, the c_{ij} are not invariant to the relative amounts of the inputs, though their signs are. As the obverses of the partial elasticities of substitution, the c_{ij} can also be defined from the cost function (from the system of equations (2.29) and (2.31)) with much more complexity as

$$c_{ij} = \frac{CG_{ij}}{w_i w_j |G|} , \tag{2.36}$$

where $|G|$ is the determinant of the bordered-Hessian matrix that results from totally differentiating (2.29) and (2.31), and G_{ij} is the cofactor of g_{ij} in that matrix.[8]

Unlike the two-factor case, in which $c = 1/\sigma$, $c_{ij} \neq 1/\sigma_{ij}$. One cannot even infer the sign of the partial elasticity of complementarity from that of the partial elasticity of substitution between the same two factors. While σ_{ij} is calculated on the assumption that output is constant, calculating c_{ij} assumes marginal cost is constant. It is possible that changes in relative wages change marginal costs in such a way as to cause the equality to disappear. As an example, employers would react to an increase in the relative wage of young workers (perhaps the abolition of a subminimum wage for youths) by substituting adult female workers for youths, so that $\sigma_{ij} > 0$. An influx of adult women into the labor force could lower the relative wages of young workers, so that $c_{ij} < 0$.

multifactor case when we assume $N = 2$. Remembering that $s_L\sigma_{LL} + s_K\sigma_{KL} = 0$, $\eta_{LL} = -s_K\sigma_{KL}$. Since $s_K = 1 - s_L$, and σ_{KL} is just alternative notation for σ, the two representations are identical.

[8] Sato and Koizumi (1973, 48) derive the c_{ij} from a cost function.

Analogous to a factor-demand elasticity is

$$\frac{\partial \ln w_i}{\partial \ln X_j} = \epsilon_{ij} = s_j c_{ij} , \tag{2.37}$$

the *partial elasticity of factor price i* with respect to a change in the quantity X_j.[9] Since $\epsilon_{ii} = s_i c_{ii} < 0$, and $\Sigma_j s_j c_{ij} = 0$, $\epsilon_{ij} > 0$ for at least one input. It is possible, though, that there are factors for which $\epsilon_{ij} < 0$ for some $j \neq i$, that is, for which an exogenous increase in the quantity of input j reduces the price of input i at a constant marginal cost. For example, an influx of new immigrants into a labor market must raise the wage rate of at least one other group of workers, or increase the rate of return to capital; but it could lower the wage received by some other group of workers (presumably a group that competes for jobs with the new immigrants).

The partial elasticities of demand and of factor prices can be used to classify the relationships within pairs of factor inputs, with a terminology based on whether quantities (q) or factor prices (p) are assumed to shift exogenously. Using the ϵ_{ij}, inputs i and j are said to be *q-complements* if $\epsilon_{ij} > 0$. They are *q-substitutes* if $\epsilon_{ij} < 0$. It is possible for all input pairs (i,j) to be q-complements, but the interesting case arises when inputs in at least one pair are q-substitutes. Using the η_{ij}, inputs i and j are said to be *p-complements* if $\eta_{ij} < 0$. They are *p-substitutes* if $\eta_{ij} > 0$. It is possible for all input pairs (i,j) to be p-substitutes, but the problem is more interesting if inputs in one pair are p-complements. If there are only two inputs, they must be q-complements and p-substitutes. An increase in the price of capital must induce firms to use more labor at a constant output; an increase in the amount of capital in a market raises the productivity of labor and hence its wage.

[9] This can be derived by totally differentiating (2.29) and (2.31) under the assumption that G is linear homogeneous to obtain

$$[G] \begin{bmatrix} dY/Y \\ dw_1 \\ \vdots \\ dw_N \end{bmatrix} = \begin{bmatrix} Yd\mu \\ dX_1 \\ \vdots \\ dX_N \end{bmatrix} .$$

Solving in (2.29) for $\partial w_i/\partial X_j$ yields

$$\frac{\partial w_i}{\partial X_j} = \frac{G_{ij}}{|G|} .$$

Multiplying both numerator and denominator in this expression by $Cw_iw_jX_j$ gives (2.37).

Some examples may help demonstrate the use of these definitions. If educated and uneducated workers are p-substitutes, one may infer that a rise in the cost to employers of employing the low-wage, uneducated labor, perhaps resulting from an increase in the minimum wage, will increase the fraction of educated workers used at each level of production. These two factors may also be q-complements. If so, an increase in the relative supply of educated workers (perhaps resulting from increased awareness of the nonpecuniary benefits of acquiring a college education) will raise the relative wage of uneducated workers by making them relatively more productive.

These derivations and the specific examples illustrated below provide explicit ways of inferring the underlying production parameters that determine the relevant own- and cross-partial factor-demand elasticities, and the own- and cross-partial factor-price elasticities. If one is less interested in the formalism and simply wishes to examine demand elasticities absent the theoretical structure, one can use Shephard's lemma in (2.29) to derive

$$X_i^* = X_i^d(w_1, \ldots, w_N, Y),\ i = 1, \ldots, N, \tag{2.38}$$

a generalization of (2.9a'). The logarithm of (2.38) for a particular input k yields a reasonable loglinear form for estimation, though by ignoring the other N-1 equations in (2.38) the researcher discards substantial amounts of information that could be relevant for the factor-demand elasticities η_{ki} of interest.

A. *Multifactor Cobb-Douglas and CES Functions*

These are just logical extensions of the two-factor cases. The N-factor Cobb-Douglas cost function can be written

$$C = Y\prod_i w_i^{\alpha_i},\ \Sigma\alpha_i = 1. \tag{2.39}$$

Each $\sigma_{ij} = 1$ (as can be seen by applying (2.33) to (2.39)), making this function quite uninteresting in applications where one wishes to discover the extent of p-substitutability (measure cross-price elasticities) or examine how substitution between X_i and X_j is affected by the amount of X_K used. That $c_{ij} = 1$ can be readily derived from a generalization of the argument in (2.16)–(2.18). The only reason for estimating an N-factor Cobb-Douglas function is to discover the shares of output accounted for by each of the inputs. The production parameters can then be estimated using data on costs, output, and factor prices in an equation that takes logarithms of both sides of (2.39).

The N-factor CES production function is

$$Y = \left[\Sigma\beta_i X_i^\rho\right]^{1/\rho},\ \Sigma\beta_i = 1. \tag{2.40}$$

As with the N-factor Cobb-Douglas function, the technological parameters are identical for all pairs of inputs and thus not very interesting:

$$c_{ij} = 1 - \rho \text{ for all } i \neq j.$$

A slightly more interesting case is Sato's (1967) two-level CES function containing M groups of inputs, each of which contains N_i individual inputs:

$$Y = \left\{\left[\sum_1^{N_1}\alpha_i X_i^{\rho_1}\right]^{\nu/\rho_1} + \cdots + \left[\sum_{N_{M-1}}^{N_M}\alpha_k X_k^{\rho_M}\right]^{\nu/\rho_M}\right\}^{1/\nu},\ \sum_1^{N_M}\alpha_i = 1, \tag{2.41}$$

where ν and the ρ_m are parameters to be estimated. Equation (2.41) is the same as (2.40), except that groups of factors aggregated by CES subfunctions are themselves aggregated by a CES function with the parameter ν. Let $\sigma_\nu = 1/[1 - \nu]$, and $\sigma_m = 1/[1 - \rho_m]$. Then for pairs of inputs i,j in different subgroups, $\sigma_{ij} = \sigma_\nu$. For factors i and j within the same subaggregate m,

$$\sigma_{ij} = \sigma_\nu + \frac{1}{s_m}[\sigma_m - \sigma_\nu],\ m = 1, \ldots, M,$$

where s_m is the share of inputs in group m in total cost. Each level of this production function can be estimated, for example, using a version of (2.21), to show substitution within each pair of inputs and then between each pair of aggregated inputs. The multilevel CES form is still quite restrictive, though. It retains the assumption that the ease of substitution is the same between all pairs of factors not in the same subgroup. It also imposes separability—substitution within a subgroup is unaffected by the amount of inputs from other subgroups; and most seriously, it requires the researcher to choose how to group inputs into particular subgroups.

B. *Generalized Leontief*

The cost function, an expanded version of (2.23), is

$$C = Y\sum_i\sum_j a_{ij}w_i^{.5}\, w_j^{.5},\ a_{ij} = a_{ji}. \tag{2.42}$$

The technological parameters are estimated from the system of linear equations:

$$\frac{X_i}{Y} = a_{ii} + \sum_{j\neq i} a_{ij}\left[\frac{w_j}{w_i}\right]^{.5},\ i = 1, \ldots, N. \tag{2.43}$$

This approach has the virtue for studies of the demand for different types of labor that one can easily add nonwage variables that might affect the number of workers in a labor market or industry. The partial elasticities of substitution are

$$\sigma_{ij} = \frac{a_{ij}[w_i w_j]^{.5}}{2s_i s_j},$$

and

$$\sigma_{ii} = \frac{w_i}{2s_i^2}\left[a_{ii} - \frac{s_i}{w_i}\right].$$

To calculate the σ_{ij}, only those parameters that involve factors i and j are used.[10] A production function similar to (2.42) yields an analogue to (2.43) that has the ratio of a factor price to cost on the left side of each equation and ratios of quantities on the right side. With this version one can derive the c_{ij} using only the parameters involving the terms in X_i and X_j. It is particularly useful (and is used in most of the empirical work based on this function) if the only inputs are various types of labor, so the production system has wage rates as functions of relative labor inputs.

C. *Translog and CES-Translog*

In general the translog cost function is

$$ln\ C = ln\ Y + a_0 + \sum_i a_i ln\ w_i + .5\sum_i \sum_j b_{ij} ln\ w_i ln\ w_j, \qquad (2.44)$$

with

$$\sum_i a_i = 1;\ b_{ij} = b_{ji};\ \sum_i b_{ij} = 0, \text{ for all } j.$$

The first and third equalities in (2.44) result from the assumption that C is linear homogeneous in the w_i (proportionate increases in the w_i raise costs proportionately); the second assumption stems from the requirement on the cost function (2.29) that $g_{ij} = g_{ji}$. With some manipulation one can derive a set of share equations that are linear in the production parameters:

$$s_i = a_i + \sum_{j=1}^{N} b_{ij} ln\ w_j,\ i = 1, \ldots, N.^{11} \qquad (2.45)$$

[10] To derive σ_{ij}, perform the required differentiation, remember that $g_i = X_i$, and that the s_i equal the share of X_i in cost and revenue.

[11] By Shephard's lemma in (2.44)

$$\frac{\partial ln\ C}{\partial ln\ w_i} = \frac{X_i w_i}{C} = s_i,\ i = 1, \ldots, N,$$

where both sides of the factor demand equation have been multiplied by w_i/C, and we have assumed factors receive their marginal products. Differentiating yields (2.45).

In this system the partial elasticities of substitution are

$$\sigma_{ij} = \frac{b_{ij} + s_i s_j}{s_i s_j},\ i \neq j; \qquad (2.46a)$$

and

$$\sigma_{ii} = \frac{b_{ii} + s_i^2 - s_i}{s_i^2}. \qquad (2.46b)$$

The σ_{ij} can also be calculated from a translog production specification, using (2.32) and thus the determinant of what could be a large matrix. To derive the c_{ij} easily a production function analogous to (2.44) can be specified and manipulated to yield share equations like (2.45), but with terms in the logarithms of the input quantities on the right-hand sides. The definitions of the c_{ij} are identical to (2.46), except the parameters are based on these alternative share equations.

The multifactor CES-translog representation departs from the translog in the same manner as in the two-factor case by replacing $\sum_i a_i ln\ w_i$ in the cost function (2.44) with

$$ln\left[\sum_i a_i w_i^{1-\sigma}\right]^{1/1-\sigma}.$$

In the ith share equation the a_i is replaced with

$$\frac{a_i w_i^{1-\sigma}}{\sum_j a_j w_j^{1-\sigma}}. \qquad (2.47)$$

Though the system of share equations becomes nonlinear once these terms are included, it can be estimated by nonlinear techniques. The gain is greater generality in the estimates of the partial elasticities of substitution.[12]

As in the two-factor case, the assumption of homotheticity can be readily relaxed by specifying that unit costs, C/Y, depend on Y. In the translog case this means adding terms $\delta_j ln\ Y \cdot ln\ w_j$, $j = 1, \ldots, N$, with $\Sigma\delta_j = 0$, to (2.44). Each share equation (2.45) then includes a term in $\delta_i ln\ Y$. This extension of the translog function is readily suited for empirical work, though it has not been very widely used by labor economists.

Throughout I have strictly divided the discussion of the two polar cases implicit in Figures 2.1a and 2.1b. Either factor prices have been

[12] The partial elasticity of substitution between factors i and k is

$$\sigma_{ik} = \left\{(\sigma - 1)\left[\frac{a_i w_i^{1-\sigma}}{\sum_j a_j w_j^{1-\sigma}}\right]\left[\frac{a_k w_k^{1-\sigma}}{\sum_j a_j w_j^{1-\sigma}}\right] + b_{ik} + s_i s_k\right\} \Big/ s_i s_k.$$

assumed to be exogenous (the production-function approach), or factor quantities have been assumed to be given (the cost-function approach). In a variety of cases one might, for example, wish to calculate the effect of an exogenous increase in the size of one group of workers on the wages of other groups of workers and on the numbers of workers employed in still other groups whose wages are rigid.

As an example, consider a world in which employment of factors $i = 2, \ldots, N$ is exogenous at X_i^*, but the wage of workers of type 1 is fixed at w_1^*. What is the effect of an exogenous increase in the number of workers of type k, ΔX_k^*, on employment of type 1 workers and on the wages of the other N-2 types of workers? Marginal productivity conditions analogous to (2.30) are

$$w_1^* = f_1(X_1, X_2^*, \ldots, X_N^*), \tag{2.48a}$$

and

$$w_i = f_i(X_1, X_2^*, \ldots, X_N^*),\ i = 2, \ldots, N. \tag{2.48b}$$

Substantial differentiation and manipulation of (2.48) yield

$$\frac{\partial ln\ X_1}{\partial ln\ X_k^*} = -\frac{s_k c_{1k}}{s_1 c_{11}}, \tag{2.49a}$$

and

$$\frac{\partial ln\ w_i}{\partial ln\ X_k^*} = \frac{s_k[c_{ik}c_{11} - c_{1i}c_{1k}]}{c_{11}},\ i = 2, \ldots, N. \tag{2.49b}$$

If one knows the factor shares and partial elasticities of complementarity, one can multiply the elasticities in (2.49) by $\Delta X_k^*/X_k^*$ to obtain the percentage changes in employment of the one factor and the factor prices of the others.[13]

V. Labor Demand in Nonprofit Organizations

The increasing importance of government and other nonprofit organizations justifies at least a brief look at how the static theory of employment demand needs to be modified when one moves away from profit maximization. Consider first a firm that is just like the archetype of the previous three sections, but that does not seek to maximize profits. It does minimize costs, and it does hire the same types of productive factors as the profit-maximizing firm. Its goals could be

[13] This example is modeled on Grant and Hamermesh (1981). The general case in which some input prices are rigid while others can vary freely is described by Johnson (1980).

avoiding losses (maintaining nonnegative profits) or maximizing revenue. A wide range of other goals is possible (Reder 1975), but these very straightforward ones probably characterize a lot of nonprofit industry. Whether they describe government employers, who may get intrinsic rewards from hiring workers of different groups, is unclear. In the case of government the applicability of the results of this chapter has to rest on their being a good approximation to what government employers do.

If the nonprofit firm is a cost minimizer, its cost function is qualitatively the same as (2.8). Under the same assumption (that production is linear homogeneous in the inputs) labor demand can be described by (2.9a), with demand depending only on factor prices. More generally, as long as the production function is homothetic, we can write labor demand as

$$L^* = C_2(C^1(Y) \cdot C_w^2(w,r)), \tag{2.50}$$

so that any effects of changes in factor prices are separable from those of changes in output. Thus, in this case η_{LL} and η_{LK} are the same as before (given the description of technology in the particular production function).

The η'_{LL} and η'_{LK} will differ from the profit-maximizing case, because scale effects will not be the same. In the competitive case it is difficult to examine what will occur. But between profit-maximizing and revenue-maximizing noncompetitive firms that have the same production functions, the former will take account of changes in revenue and increases in cost as it expands, while the latter will only consider revenue. The latter will thus raise output more in response to a given drop in an input price, so that scale effects will be greater so long as average costs eventually increase.

The situation is more complicated still among cooperative firms, where the goal may be to maximize net revenue per cooperative member. The stylized ideal here is the labor-managed firm (Vanek 1970), which is assumed to maximize revenue minus other input costs per worker:

$$\text{Net revenue} = \frac{PF(X_1, \ldots, X_N, L) - \Sigma X_i}{L}. \tag{2.51}$$

In the multifactor case the results of Section IV are unchanged if production is separable in cooperative members L, for then L can be treated exactly as entrepreneurs are in the case of the profit-maximizing firm: They maximize an analogue to profits in full awareness that the size of their own input does not affect substitution possibilities among other inputs.

The assumption of separability of labor (co-op members) from other inputs is not very credible; but dropping it makes it hard to draw many inferences about factor demand. Assume the co-op is small, and let the price of its product be one. Then, following Meade (1972), consider the simple case in which it hires capital services to work with its members and maximizes:

$$\text{Net revenue} = \frac{F(L,K) - rK}{L}. \tag{2.51'}$$

In this case the net-revenue-maximizing conditions are

$$F_K = r, \tag{2.52a}$$

and

$$F_L = \frac{F(L,K) - rK}{L}. \tag{2.52b}$$

Condition (2.52a) is identical to (2.5c): The co-op hires capital services until their marginal product equals their rental price. It expands its membership until the marginal product of another member equals the net revenue that each member can draw from the co-op. The only exogenous factor price is r.

None of the long-run factor-demand elasticities (with respect to r) is the same as in the profit-maximizing case. With output constant the response to an increase in r is affected by the change in the numerator in (2.52b). If output can vary, the responses are still more complex, with even the directions of the effects on K and L being indeterminate.[14] Other than being sure that at a constant output an increase in r will lower K and raise L, we cannot say very much about labor demand in this case.

VI. The Distinction between Workers and Hours, and the Cost of Labor

Throughout the previous sections we assumed that all workers exerted the same amount of effort in the workplace and, even more

[14] In this situation

$$\frac{dK}{dr} = \frac{[-LF_{LL} - KF_{LK}]}{\Delta},$$

and

$$\frac{dL}{dr} = \frac{[LF_{LK} + KF_{KK}]}{\Delta},$$

where $\Delta = F_{LK}^2 - F_{LL}F_{KK} < 0$. The directions of the responses to a change in r depend on the shape of F and on the capital-labor ratio.

important, that this amount was not subject to choice by either the worker or the employer. Indeed, the term "labor" was never strictly defined. Once we wish to talk of hours, more precision is necessary. Therefore throughout the rest of the volume the term "labor" denotes a particular subaggregate of workers, or the total input of time from that subaggregate. "Worker-hours" denotes the product of "employment," the number of employees in the group, and their "hours," the amount of time they work per period. A similar lack of precision has attached to the "wage," which has implicitly thus far been just the "price of labor" in the long run. That too no longer suffices, so that in discussing choices between workers and hours here, and in analyzing adjustment in Chapter 6, I distinguish among various components of labor costs.

Essentially we have implicitly assumed for each type of labor i that

$$L_i = E_i H_i, \tag{2.53}$$

where E is employment, and H is the hours they work per time period. The assumption that effective labor input is multiplicative in employment and hours masked any need to consider differences in the prices of these two possible ways of altering the input of labor. The assumption is clearly unrealistic, particularly along the hours dimension: Doubling weekly hours from 60 to 120 will probably not double the amount of effective labor. Indeed, whether effective labor would even increase is questionable. Given the absurdity of the behavioral assumption implicit in (2.53), it makes sense to see what additional insights can be gained from generalizing about how workers and hours are combined.

Moving beyond (2.53) means that we view employers as facing interesting choices about how much labor to employ at the *extensive margin* of additional employment and the *intensive margin* of planning for a work force that can expend more or fewer hours per time period. We measure the intensity of production solely by H, therefore ignoring the realistic possibility that effort per hour worked can vary. Perhaps most important, we do not deal with variations in hours worked in response to short-run changes in derived demand or with the time path of hours between long-run equilibria. These questions have provoked much of the study of the demand for hours, and they are dealt with in Part II. This means, though, that we cannot assume that the firm's capital stock is fixed (though it may be reasonable to assume that the demand for workers and hours is separable from the demand for capital services).

Hours H are measured per time period; but per *what* time period—day, week, month, year, lifetime? The convention is to measure H as

TABLE 2.1
Weekly Hours of Work by Industry, United States, 1990

Industry	Hours
Mining	44.0
Construction	38.2
Manufacturing	40.8
Transportation and Public Utilities	38.9
Wholesale Trade	38.1
Retail Trade	28.8
Finance, Insurance, and Real Estate	35.8
Services	32.6

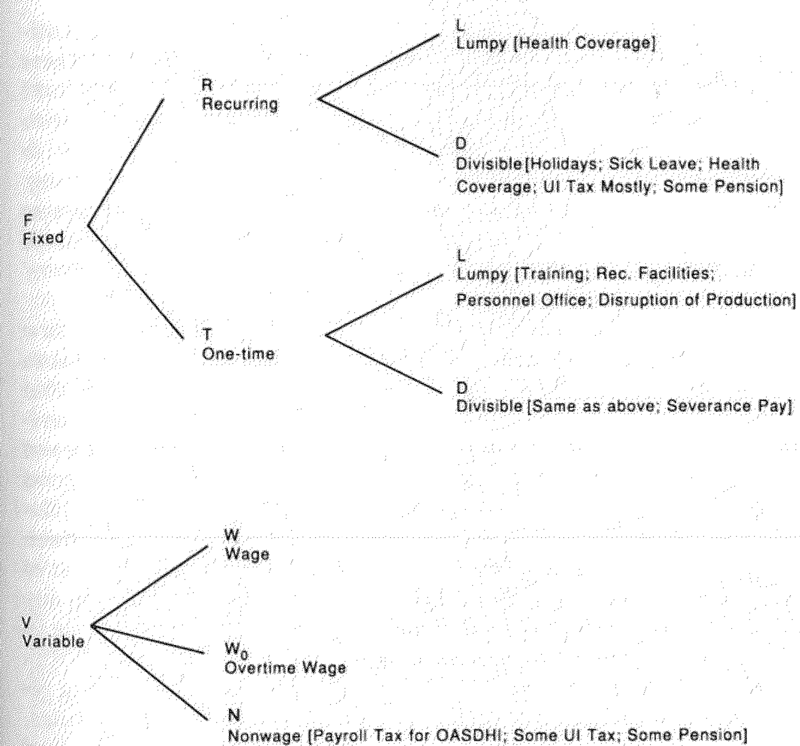
Source: Employment and Earnings, March 1991, table C.2.

hours per week, and I stick with that convention. Nonetheless, it is important to recognize that scheduled weekly hours need not equal 1/52 of scheduled annual hours and, more important, that there are differences in employers' costs of varying hours per week and weeks per year that will affect decisions about these dimensions of the intensity with which employees are used.

That employers do have sufficient scope for substituting hours for workers is demonstrated by the very sharp differences in weekly hours even among the broadly defined industries shown in Table 2.1. (See also Lilien and Hall 1986.) The data suggest that technology differs among industries in ways that dictate differences in work intensity, that there are interindustry differences in the relative costs of workers and hours, or some combination of both explanations.

The designation of all labor costs in Sections II–IV as a wage, w, must be abandoned when we move to examining choices about workers and hours. Regrettably, there is no clear-cut typology for labor costs, though some have been suggested (Rosen 1968; Hart 1984). Here, I concentrate on developing a new description of costs that seems parsimonious, yet is sufficiently exhaustive to provide all the distinctions necessary for the analysis here and in subsequent chapters.

The main distinction is between costs that vary with E, *fixed costs*, measured throughout this section on a per-worker basis as F, and those that vary with hours, *variable costs* V. This distinction is the basis for the two major branches of the typology shown in Figure 2.2; yet attempts to pigeonhole specific aspects of labor costs even at this low level of distinction require care. In the United States the payroll tax that finances state unemployment insurance benefits has a ceiling on the taxable annual wages that averaged (in 1991) roughly $9,000.



2.2 Typology of Labor Costs

This means that for most workers this tax represented a fixed cost, since additional hours worked did not raise the employer's tax liability. For very low-wage workers, though, it was a variable cost, for extra hours raised the tax bill.[15] Near the other extreme in the United States is the payroll tax for Old Age Survivors' and Disability Insurance (OASDI), which has a very high maximum taxable earnings ($53,400 in 1991). On all but the highest-wage workers this tax is a variable cost to employers; but in considering whether to employ high-wage workers and how intensively to use them, it is a fixed cost.

Within the category of fixed costs a useful distinction exists between *recurring* costs, R, and those that are incurred at *one time*, T,

[15] The UI tax is also a good example of a labor cost that differs along the dimensions of weekly hours and weeks per year.

usually when the worker is hired. Employer-provided health insurance is a good example of recurring fixed costs. Covered workers generate the same premium cost regardless of their work hours, and the premium is paid every month. Pension costs do vary in part with hours worked, but under many plans the variation is not linear or even constant, so that part of these costs is also fixed and recurring. One-time fixed costs include the reduction in output that occurs as inexperienced workers are trained up to full capacity, the costs of operating a personnel office, and at the other end of some workers' tenure, any severance pay.

A still finer distinction exists within both recurring and one-time costs between those fixed costs that are *lumpy*—that are invariant to the number of workers—and those that are *divisible* and thus that vary with employment. Among recurring costs much of health coverage can be viewed as lumpy: There clearly are some economies of scale in providing such coverage. Holidays and sick leave are good examples of divisible recurring costs, since the extra costs vary linearly with the size of the work force. Making the distinction between lumpy and divisible onetime costs is more difficult, because most of these, and particularly the direct and indirect costs of training, are hard to observe. Nonetheless, there are some economies of scale in hiring and in the activities of the personnel office that result from spreading out lumpy costs. Some government-imposed reporting requirements, including some produced by affirmative-action rules, generate costs only if a hire occurs, and the same forms can be used for each hire. The distinction between lumpy and divisible costs is not used here; but it is important for examining employment dynamics in Part II.

The distinctions among variable costs are less well articulated, since such costs are less diverse. It is useful, though, to distinguish among standard and overtime wages, and between wages and such nonwage costs as parts of pension, OASDI, and unemployment insurance payments. These distinctions necessitate examining how the price of an additional hour per week varies with hours.

All of the one-time costs are identical analytically to the costs of capital: They are incurred only once during the tenure of the worker in the plant, and they generate per-period costs equal to $r+q$, where q is the quit rate.[16] The typical firm thus faces fixed costs per period of

$$EF = E\{R + [r+q]T\}.$$

[16] If workers quit at a rate q per period, the initial costs they generate can be recovered with a sinking fund requiring periodic payments of q cents per dollar.

The borrowing rate r is exogenous to the firm's choice between workers and hours; but the quit rate may well not be, as higher onetime fixed costs may lead firms to alter their wage policy in order to reduce turnover costs (Hamermesh and Goldfarb 1970; Pencavel 1972). Though no doubt correct, this extension is not central to the results on the worker-hours distinction. Also, while empirical work requires distinguishing between R and T, there is no gain to doing so in the theoretical exposition. In sum, throughout the rest of this chapter I parameterize fixed costs per worker as F, recognizing that any results on the effects of higher F can reflect higher values of R, r, q, or T.

Conventional analysis of the choice between workers and hours treats the various components of labor income, in particular the wage rate and any premium for overtime work, as exogenous to the firm. I analyze briefly that type of model at the end of this section. But that view is quite inconsistent with the huge corpus of literature on labor supply. If typical workers are asked to work additional hours, they will require a higher wage rate to do so. Indeed, if workers have identical tastes, the firm will face a wage

$$w = w(H),\ w' > 0, \tag{2.54}$$

with w' equaling the marginal rate of substitution of income for leisure in the typical worker's utility function at the equilibrium wage rate and weekly hours. The theory of labor supply suggests we should not treat the wage as invariant to employers' hours decisions. That in turn implies that imposed changes in one aspect of the wage-hours package will affect the other through demand *and* supply (Trejo 1991). That view conditions the discussion of the effects of overtime laws in Chapter 5. In the meantime, the general model recognizes this interdependence by using (2.54) to specify the wage-hours relationship.

The typical firm is constrained by supply to pay workers a wage $w^* = w(H^*)$ determined by (2.54). I assume the firm is small enough that it faces an infinitely elastic supply of potential employees; but each potential employee has an upward-sloping reservation wage that must be paid to retain the worker's services at H^* per week. Coupled with the other assumptions, this means that the firm's labor costs are

$$\text{Labor cost} = EHw(H) + EF. \tag{2.55}$$

The cost of capital services is, as before, rK per time period.

In much of the discussion I assume that choices about workers and hours are separable from capital. This assumption is clearly not always correct: On an assembly line capital and employment may be highly p-complementary, while they may be jointly p-substitutable against hours (and the rate of utilization of the capital). Nonetheless, the initial assumption of separability may be valid and makes the analysis easier.

The second issue is the shape of the labor aggregator:

$$L = L(E,H).$$

In the firm's output maximization subject to the constraint that labor cost is C^0, it makes sense that $\frac{\partial H}{\partial C^0} = 0$, that is, that the firm's optimal hours be independent of scale. There is no evidence that weekly hours of full-time workers at General Motors differ substantially from hours of workers at the local steel fabricator. This assumption requires that

$$L = \phi_1(E)\phi_2(H),\ \phi_i' > 0.$$

Ehrenberg (1971a) has suggested a specific example of this function that is particularly easy to use:

$$L = aE^b g(H),$$

$a, b > 0$, and $g' > 0$. These functions are clearly quite restrictive, but the restriction seems consistent with the observation that weekly hours are basically invariant with scale.

To make the exposition easier, I present the firm's problem as one of maximizing L subject to the constraint that labor cost equals C_0. This yields, after some manipulation,

$$\frac{L_E}{L_H} = \frac{wH + F}{wE[1 + \epsilon]}, \tag{2.56}$$

where ϵ is the elasticity of wages with respect to hours. As is standard, the ratio of the marginal products of employees and hours is set equal to the ratio of their relative prices when the firm maximizes its labor input with a cost constraint.

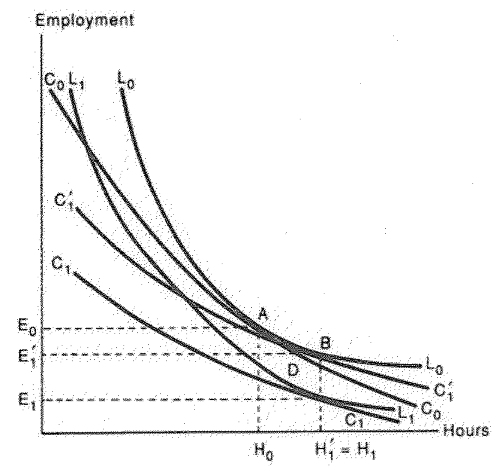
Notice that the exogenous variables facing the firm are the fixed costs, F, and the (constant) elasticity of the wage rate, ϵ. I assume that the former is determined by technology and the latter by workers' preferences. From (2.56) it is clear that an increase in F raises the price of employees, while an increase in ϵ raises the price of hours. Since the marginal products of each are positive and decreasing, the demand functions are

$$E^* = E(\overset{-}{F}, \overset{+}{\epsilon}, \overset{+}{C^0}), \tag{2.57a}$$

and

$$H^* = H(\overset{+}{F}, \overset{-}{\epsilon}, \overset{0}{C^0}), \tag{2.57b}$$

where the superior signs denote the effects on employment or hours of increasing the parameter in question. These are the fundamental



2.3 Substitution and Scale Effects in the Worker-Hours Choice

results of the discussion in this section. If choices of workers and hours are separable from capital, an exogenous increase in fixed employment costs reduces the ratio of employment to hours at a given scale. An exogenous increase in the elasticity of wages with respect to hours raises this ratio. An increase in scale will, by assumption, affect only employment in the long run.

These substitution effects, and the scale effects that also result, can be illustrated by Figure 2.3. The initial labor input is L_0, consisting of employment E_0 and hours H_0. The labor isocost is C_0, which is convex as long as w' is not too positive.[17] Consider the effect of an increase in fixed employment costs. This shifts the isocost to C_1, both lower and flatter than C_0. The firm moves from A to D in the diagram, reducing employment to E_1 and increasing hours to H_1. The substitution effect along the isolabor curve L_0 to its tangency with the isocost C_1' that is parallel to C_1 is AB, an unambiguous increase in hours and cut in employment. In addition, the increase in fixed costs increases the cost of each worker-hours combination, so that the scale of operations is reduced. This leads to the scale effect BD. By assumption there will be no scale effect on hours, consistent with observation, though it is possible theoretically that the scale effect could change hours. It is certain that employment will be reduced and that EH, worker-hours, are reduced by this increase in fixed costs.

[17] The degree of convexity of C_0 must be less than that of the isolabor curve L_0 to obtain an internal maximum.

It is worth noting how changes in the parameters F and ϵ affect the equilibrium wage rate (remembering that in this general model the wage is determined by the interaction of the typical worker's labor-leisure choice and the firm's decision about employment and hours). The rise in equilibrium hours that is produced by an increase in fixed costs must in this model be accompanied by an increase in the average wage. (Otherwise, the supply of hours to the firm would not be forthcoming.) This conclusion is unaffected by any scale effects, since we assumed that scale effects have no impact on equilibrium hours.

Thus far I have assumed that workers and hours can be aggregated into the input "labor," so that choices about employment and hours are separable from choices about labor and capital inputs. If this is not so, most of the conclusions about the directions of the effects of higher fixed costs, or a higher wage-hours elasticity, no longer hold. Only the result that $\frac{\partial E^*}{\partial F} < 0$ is still valid (Hart 1984, 77–78); the effect of F on H^* is ambiguous. The reasons for these results can be seen by referring back to the discussion in Section IV. With more than two factors of production the increase in fixed costs, which is an increase in the price of E only, produces the usual negative own-price effect on E. Without knowing the relative p-substitutability or complementarity among the inputs, we cannot generally infer the impact on H.

The impacts of an increase in ϵ in (2.57a) and (2.57b) both become ambiguous once the separability of capital from labor is no longer assumed. The reason is that a higher ϵ represents an increase in two prices, those of employment and hours, relative to the price of capital services. Thus there is no unambiguous own-price effect. It is likely that the demand for hours falls, but if hours and capital are sufficiently relatively p-complementary compared to hours and workers, the demand for hours will rise. The effect on employment is ambiguous, since while its price falls relative to that of hours, it rises relative to the price of capital services.

Most of the theoretical work on the employment-hours decision has not used the general cost specification in (2.55). Instead, researchers (Rosen 1968, 1978; Ehrenberg 1971a; Hart 1984) have specified costs as

$$C^0 = EwH_s + Ew[H - H_s][1 + p] + EF, \qquad (2.58)$$

where w is the exogenous straight-time wage, p is the overtime premium, and H_s is standard hours, above which additional hours must be paid at the overtime rate. This respecification of labor costs changes the isocosts in Figure 2.3 by introducing a kink at H_s, with the isocost having a steeper slope to the right of H_s.

Assuming that all firms use some overtime, and redefining ϵ to be the elasticity of variable labor costs (the elasticity of all the terms in (2.58) except EF) with respect to H, yields

$$\epsilon = \frac{H[1 + p]}{H[1 + p] - pH_s} \geq 1,$$

with a strict inequality as long as the overtime premium is positive. This respecification of variable labor costs makes little sense as a model of labor-market equilibrium, since it completely ignores how workers' supply behavior might be affected by changes in H occasioned by altering p or F. Since it has been widely used, though, it is worth examining the implications for H^* and E^* of changes in the parameters of C^0.

The general results carry through this specific model in the case in which capital services are separable from labor. In particular, notice that a higher overtime premium raises the hours elasticity of labor costs. This means that an increase in p induces substitution away from hours and toward employees. A reduction in standard hours reduces the labor-cost elasticity and thus produces the opposite effects on employment and hours. Finally, as before, higher fixed costs reduce the ratio of employees to hours.

All of these inferences are made under the assumption that there are no scale effects on employment. (That there is none on hours is the maintained assumption throughout this and the next section.) All three changes—a higher overtime premium, lower standard hours, and higher fixed costs—produce negative scale effects on the demand for employment. Thus, the negative substitution effects on employment of lower standard hours and higher fixed costs are exacerbated by the scale effect; and the positive substitution effect of higher overtime premia is mitigated, and perhaps even reversed, by the negative scale effect it induces. The negative scale effect generated by lower standard hours is likely to be especially large, since it produces a large increase in the cost of inframarginal labor services.

These conclusions can only be drawn readily under the assumption that the typical firm works its homogeneous labor force some amount of overtime hours. In a more general model, in which some firms are in an overtime regime while others are not, a rise in fixed employment costs causes some firms to shift out of the straight-time regime and to use overtime hours. The possibility that firms shift regimes in response to changes in the cost parameters complicates the analysis. However, the regime shifts are in the same directions as the marginal changes made by firms that continue to use overtime. Thus, the conclusions about the directions of the substitution effects and the total impacts still hold.

The same cannot be concluded about the responses to a drop in standard hours. That change will cause some firms to shift to the straight-time regime, implying that total hours in those firms are reduced. This shift may be sufficient to outweigh the positive impact on hours among firms that remain in the overtime regime, so that the substitution effect on hours becomes ambiguous. Coupled with the scale effect on employment, though, it means that dropping the assumption of identical firms reduces and may even reverse the overall negative impact on employment.

As in the general model, when the separability of capital services from labor is not assumed, nearly all predictions about factor substitution become ambiguous. Unless one specifies the possibilities for substitution among employees, hours, and capital services, only the result that higher fixed costs reduce employment still holds. The reasons are the same as before: Most of the changes can operate on two margins, between workers and hours, and between workers (or hours) and capital services.

As noted above, workers' supply responses, which require that a higher wage rate must be paid to elicit additional weekly hours, make it unlikely that the wage rate remains fixed when the firm's cost parameters change. Assume the firm must offer a wage-hours package that maintains the typical worker's utility at

$$U(wH_s + w[1 + p][H - H_s], \bar{T} - H) = \bar{U}, \qquad (2.59)$$

where $\bar{T}$ are the total weekly hours available to the typical worker, $\bar{U}$ is the utility level available in other firms, and $U_i > 0$, $U_{ii} < 0$. Then if capital services and labor are separable, the negative substitution effect of a higher overtime premium on hours raises the second argument in U. As long as employment-hours substitution is not very large, the first argument will also increase. The firm can then reduce the wage and still attract workers, for it can maintain $U = \bar{U}$. At the very least this means that some part of the negative scale effect on employment generated by the increased overtime premium is eliminated by the action of the market. In this broader, labor-market view of the worker-hours decision a higher overtime premium results in lower straight-time wages, and because of that the reduction in hours produced through the substitution effect will be smaller.

Some ambiguity exists with fixed costs. On the demand side higher fixed costs result in higher weekly hours, thus reducing utility through the second argument of U; but the increase in hours raises income, thus raising utility in (2.59). Presumably, following our assumption of an upward-sloping labor supply schedule earlier in this section, the wage must increase with hours worked to hold utility at $\bar{U}$. Since an increased wage raises the cost of an hour of labor relative to that of another employee, this additional impact means that a higher overtime premium will reduce the extent of substitution toward workers and produce still larger negative scale effects on employment than in the standard model. It is possible that the wage adjustment will be sufficient so that the only impact of higher fixed costs is the negative scale effect on employment.

As we saw, a reduction in H_s generates substitution toward increased hours and away from workers and produces a negative scale effect on employment. Whether this conclusion is modified when a labor-market view is taken is ambiguous. The higher hours reduce the second argument of U in (2.59), thus reducing utility; but the reduction in H_s raises the first argument, thus increasing utility. If these labor-leisure choices are important, it is possible that the market will generate a change in the wage rate that could alter the conclusions about scale effects.

VII. The Demand for Hours in a Heterogeneous Work Force

If we abandon the assumption that workers are identical in production, the theory of labor demand offers very few concrete results. Those that can be drawn are illustrated by a model with only two types of workers, an assumption I maintain in this section. The workers are assumed to be distinguished by their skills, and this distinction means that each group of workers has its own wage-hours relation, with $w_1(H_1) < w_2(H_2)$, so that by assumption Type 2 workers are more skilled than Type 1 workers. Labor costs are

$$C^0 = \sum_{i=1}^{2} [E_i H_i w_i(H_i) + E_i F_i]. \qquad (2.60)$$

The production technology can be written like (2.28):

$$Y = f(E_1, H_1, E_2, H_2, K), \qquad (2.61)$$

a form that is so general that, without further specification, we cannot infer anything beyond the conclusions of Section IV. Therefore, consider the first and probably most interesting issue: Is there substitution between E_i and H_j independent of E_j? If not, can each type of labor be aggregated so that (2.61) can be rewritten as

$$Y = f(L_1(E_1, H_1), L_2(E_2, H_2), K), \qquad (2.61')$$

so the issue becomes one of first examining employment-hours substitution within each type of labor, as in Section VI, and then examining substitution among the three inputs L_1, L_2, and K, as in Section IV? If so, all the conclusions of Section VI apply to each type of labor. For example, higher fixed costs of Type 1 labor induce substitution toward hours of Type 1 workers and away from employing Type 1 workers, as does a lower wage-hours elasticity among these workers.

It is difficult to believe that the technology in (2.61′) describes reality. Increased hours of unskilled workers may increase the productivity of each hour worked by skilled workers. To some extent the weekly work schedule functions as a public good within the plant (Stafford 1980), so that H_1 and H_2 cannot be separable. For example, if semiskilled workers spend more hours on the assembly line, the firm may benefit by increasing the daily or weekly hours of the skilled machine repairers who keep the equipment in satisfactory operating condition.

This counterargument suggests the interesting specific possibility that workers and hours are separable:

$$Y = f(E(E_1, E_2), H(H_1, H_2), K). \qquad (2.61'')$$

If this alternative describes production well, the analysis of Section VI applies mutatis mutandis to the firm's choice between its total employment aggregate and the aggregate of hours. Combined with a discussion of how E_1 and E_2 are aggregated to generate employment (and similarly for the aggregation of hours), that analysis would provide helpful insights into the effects of changes in fixed costs or wage-hours elasticities of each type of labor.

Let the two types of labor be production workers and managers, and assume that the cost of hiring or training production workers falls, producing an increase in their employment. The increased ratio of production workers to managers raises the productivity of *both* the hours and the number of managers. This shows that employment and hours are not generally separable, and that (2.61″) will also not always be a good description of production.

These examples suggest that there will be substitution between E_i and H_j independent of E_j. It means that in general nothing can be concluded about the effects of, say, higher F_1 on the demand for hours of Type 1 workers, or on hours or employment of Type 2 labor. The own-price effects of wage-hour elasticities still hold if labor is separable from capital services; and the own-price effects of fixed costs hold if even it is not. But the direction of the substitution effects on the other components of labor input cannot be determined generally.

Based on the evidence that weekly hours do not vary systematically with firm size, I assumed in Section VI that there are no scale

effects on hours. This assumption is presumably equally valid (or invalid) when labor is disaggregated into several types. Whether the demand for hours is homothetic in the intensity with which each type of labor is used is less clear. If the firm can vary continuously the amount and utilization rate of each type of worker, there is no reason a priori to reject homotheticity. Nonetheless, there are no obvious facts that allow one to assume that the demand for hours is homothetic.

Throughout this section I have assumed heterogeneity exists only along the dimension of workers' skills. Workers' tastes for weekly hours were assumed to be identical, both here and in Section VI. What if they are not, and instead there is a continuum of workers arrayed by their marginal rates of substitution of income for leisure at each wage rate? In that case the market will generate an upward-sloping locus of wage-weekly hours equilibria (see Rosen 1974, 1978). The typical firm will still see itself as confronting the same wage-hours function as in Section VI, and the results derived there will still be valid. The only difference is that, rather than only generating changes in hours worked by the typical employee, parametric changes in labor costs alter the sorting of workers among firms.

VIII. Summary, and Prospects for the Theory of Static Labor Demand

The neoclassical theory of static labor demand has provided a framework and a number of specific predictions for studying how changes in exogenous factor prices and their components affect the relative and absolute amounts of labor inputs and their components. It has also generated predictions about how changes in exogenous factor quantities affect the relative and absolute wage levels of different groups of workers. The major conclusions are:

1. The effect of an increase in the wage of one group of workers on the amount of their labor demanded is negative. This negative response consists of a negative effect at a constant level of output, and a negative scale effect. An exogenous increase in the quantity of one type of labor available in the labor market produces a negative effect on those workers' wage rates. This response consists of a negative effect at a constant rate of marginal cost and a negative cost effect.

2. If there are only two inputs—say, labor and capital services—an increase in the wage raises the amount of capital services demanded at each output level; but the negative scale effect may still result in an overall decline in the firm's demand for capital services. Obversely, an increase in the supply of labor to the market raises the

return on capital services at a fixed cost of output, though the cost effect may produce an overall decline in the return to capital. If there are several inputs, perhaps several groups of workers and capital services, at least one input must see its employment increase at a fixed level of output if the wage of another group of workers rises. Similarly, the wage of at least one group of workers, or the return to capital, must rise at a given marginal cost if the available number of workers of another group increases.

3. The theory provides us with a useful framework and terminology for classifying demand relationships. Discovering whether particular pairs of inputs are p-substitutes or complements and q-complements or substitutes is helpful for evaluating the impact of a wide variety of policies, for analyzing the potential effects of changes in those policies, and for predicting how new policies will affect employment and/or wages. An increasingly elaborate superstructure of forms for estimating the underlying production relations has been built that can enable econometric research to provide estimates of these substitution relationships.

4. An increase in fixed costs of employment causes employers to alter the mix of worker-hours toward using more hours and fewer workers. An increase in the wage elasticity of additional hours (less generally, an increase in required premium wages for overtime work) produces the opposite effect. In all these cases, though, the increase in costs generates a negative scale effect that reduces total worker-hours. When combined with the substitution effect, this scale effect may be sufficiently large to cause total employment to fall when an overtime premium is increased.

The main message here is the central point of microeconomics: Price changes affect behavior. In the case of labor demand, this means that imposed increases in labor costs reduce labor demand, changes in relative wages shift relative worker-hours in the opposite direction, and relative changes in the components of labor cost alter the mix of employment and hours in the opposite direction. The changes may not be immediate. Indeed, there may not be any response if decisions about employment are lumpy, as I discuss in Chapter 4. How large the responses are is an empirical question, on which a huge amount of research has been produced (see Chapter 3). But that there is a *tendency for firms to reduce employment when wages increase and to shift relative employment toward workers who become relatively less expensive* is undeniable. Readers who are not convinced of this should close this book, as the vast body of empirical work and policy analysis is unlikely to sway them further.

Much of the development of the theory of labor demand from 1960

through 1990 was in the area of constructing functional forms for describing production technology. As Sections III and IV showed, these have been aimed at providing increasingly general methods for empirical research to infer the substitution parameters. No doubt still more complex functional forms can be invented, as the continued eruption of such forms during the 1980s (Pollak, Sickles, and Wales 1984; Considine and Mount 1984) shows. These will be useful, because they will allow further refinements of estimates of substitution relations. Despite that, extensions of this approach do not seem likely to be very fruitful in the sense of substantially broadening understanding of labor markets.

Further extensions in the long-run demand for employees and hours will undoubtedly also take place. Still more complex models of employment-hours choices along the lines of Hart (1984) can and will be built; and they may be able to provide an expanded framework that will allow careful empirical work to yield useful insights into firms' decision making and the impacts of changes in labor-market policy. Here too, though, these are extensions, amendments, and modifications rather than the basic novel research that is likely to revise and expand knowledge of the central issues in labor demand.

The most necessary theoretical work would link the demand and supply of labor. In Sections VI and VII I indicated how this approach might proceed in the context of the employment-hours choice and how it might modify conclusions about those decisions. Substantial work has proceeded since the development of contracting models in the mid-1970s (Baily 1974; Azariadis 1975). However, contracting models, which recognize how important it is to consider labor-market equilibrium rather than focus on supply and demand separately, have not been integrated into the theory of labor demand, and vice versa.

Aside from its implications for the progress of economic knowledge, this lack of communication among economists working in areas that are related can and, in this case, has resulted in serious problems when changes in policy are undertaken without paying attention to both relevant areas.[18] With that integration, and with only the data

[18] Federal legislation effective beginning in 1985 required state unemployment insurance systems in the United States to increase sharply the range of the experience-rated taxes that finance unemployment benefits. The intellectual origin of this change was the demonstration using contracting models of the disincentive effects produced by limits on tax rates under experience rating (e.g., Feldstein 1976). There was no change, though, in the very low ceiling on an employee's annual earnings that were taxable. In some states and for some employers superimposing the broader range of tax rates on the continued low base produced effective tax rates as high as 10 percent on the

now available for estimating substitution relations, much more can be learned about factor substitution. In particular, it should be possible to derive a set of estimating forms that enable one to infer the extent of factor substitution in the general context of Figure 2.1c rather than under the restrictive assumptions of perfectly elastic or inelastic supply depicted in Figures 2.1a and 2.1b. Similarly, it would enable us to understand the extent to which worker-hours outcomes are affected by both demand and supply forces.

The other area where important work on the theory of labor demand can and should proceed is what one might call the "new wage theory." This approach includes such work as the examination of efficiency wage models (e.g., Akerlof and Yellen 1986a) and the study of wage and employment outcomes that result from formal or informal bargaining over the rents generated by investments that are shared by workers and their employers (e.g., Kuhn 1988). In all this work the demand side of the models is rudimentary: A simple production function is assumed, with one type of labor at work in the typical firm. How would the conclusions of these models be altered by assuming several types of labor, especially if the extent of substitution among them and with capital is specified in ways consistent with existing empirical evidence? Obversely, bringing these models into the corpus of labor demand theory should enable us to draw better inferences about the nature of substitution among inputs.

Though the basic neoclassical theory of long-run labor demand has been well developed since the 1930s, and the major framework for applying it stems from the early 1970s, the theory need not be viewed as moribund. By integrating it into recent literatures that examine the relationships between workers and their employers, we can enhance our ability to draw useful inferences. More important from the point of view of this book, our ability to infer how employers substitute among workers and between workers and hours will also improve.

labor of their low-wage workers, as compared to rates of perhaps 2 percent on that of high-paid employees (Hamermesh 1990c). While probably reducing incentives for employment fluctuations, the modification of experience rating increased incentives to substitute high- for low-skilled employees. A dynamic imperfection was mitigated, while a static imperfection was worsened.

CHAPTER THREE

Wage, Employment, and Substitution Elasticities

I. What We Need to Infer

In Chapter 2 I developed the theory of labor demand, showing how production theory provides a framework within which empirical research can generate estimates of interesting parameters. Most important among these, and most widely studied by economists, is η_{LL}—the constant-output demand elasticity for homogeneous labor. Also included have been the total demand elasticity for homogeneous labor, η'_{LL}; the demand and factor-price elasticities for various groups of workers; partial elasticities of substitution and of complementarity between these groups, and for capital; and elasticities of substitution between workers and hours worked. In this chapter I assess critically the available estimates of these parameters and infer what we know about their magnitudes.

Empirical research on labor demand has progressed beyond being able to determine that the elasticity of substitution between labor and capital is definitively between zero and infinity (a level of knowledge that Johnson [1976, 107] claimed for it). How much further we have gone needs investigating. Achieving a consensus in this area is crucial if the theory presented in Chapter 2 is to provide an understanding of how labor markets actually work and of the likely impacts of existing and potential policies that affect them.

No single empirical study can provide definitive measures of a particular parameter. This guarantees that substantial numbers of empirical studies of the more important parameters describing labor demand will have been produced. The multiplicity of estimates imposes the burden of evaluating the design of each empirical study and, most important, of assessing whether the data allow researchers to draw the inferences they wish to make. In what follows I therefore first consider the appropriateness of various types of data, disaggregations of the labor force, and approaches to estimation for inferring labor-demand parameters. After determining the general outline of the empirical approaches that are likely to yield useful estimates, I present a detailed classification and critique of the available estimates of the parameters describing employers' demand for labor.

To: Deck, Leland[Deck.Leland@epa.gov]; Kaufman, Kathy[Kaufman.Kathy@epa.gov]; Walton, Tom[Walton.Tom@epa.gov]; Langdon, Robin[Langdon.Robin@epa.gov]; Chappell, Linda[Chappell.Linda@epa.gov]; Ferris, Ann[Ferris.Ann@epa.gov]; Evans, DavidA[Evans.DavidA@epa.gov]; Marten, Alex[Marten.Alex@epa.gov]; Shouse, Kate[Shouse.Kate@epa.gov]; Alsalam, Jameel[Alsalam.Jameel@epa.gov]; Bryson, Joe[Bryson.Joe@epa.gov]; Eschmann, Erich[Eschmann.Erich@epa.gov]; Hubbell, Bryan[Hubbell.Bryan@epa.gov]; Keaveny, Brian[Keaveny.Brian@epa.gov]; Adamantiades, Mikhail[Adamantiades.Mikhail@epa.gov]; CurryBrown, Amanda[CurryBrown.amanda@epa.gov]
Cc: Stenhouse, Jeb[Stenhouse.Jeb@epa.gov]; Weatherhead, Darryl[Weatherhead.Darryl@epa.gov]
From: Macpherson, Alex
Sent: Sat 8/1/2015 12:14:31 AM
Subject: 2 RIAs in 1 day (again!)
EO12866_CPP_Federal_Plan_2060_AS47_RIA Final_20150731.docx
EO12866_CPP_2060_AR33_RIA Final_20150731.docx

Hey team

Both (d) and FP RIAs are finished and have been sent to the project leads. They are attached and on sharepoint.

Thanks for everyones help. I'll give some updates as things progress over the weekend.

Alex

To: Marten, Alex[Marten.Alex@epa.gov]
From: Evans, DavidA
Sent: Thur 7/30/2015 5:40:46 PM
Subject: FW: Summary of Interagency Comments Under EO 12866 and 13563 -- GHG Mitigation TSD BB1
SummaryOfInteragencyCommentsUnder_EO12866_EO13563_FP.DOCX
ATT00001.htm
SummaryOfInteragencyCommentsUnder_EO12866_EO13563_CPP_RIA.DOCX
ATT00002.htm

See page 3-41 regarding request for text on differential incentives across generator. You deal with that one, I'll do w/ OBA? Or swap?

(I actually thinking I'm giving you the hard one, b/c there is so much to parse out in it). Text below.

D

While we did not model the implementation of the technology-specific rates, in practice we expect the effects of the two rates to be similar to the changes in relative operational costs faced by coal and gas plants when the relative prices of coal and gas fuels shift, as they routinely do in energy commodity markets or as occurs in existing carbon markets in California and New England where a carbon price is added to the operating costs of coal and gas generation. Power markets have routinely worked through such shifts in the costs of their inputs, and the related market dynamics are well established. For example, at times the typical long-term difference between coal and gas prices has changed significantly and durably, and the ratio of coal to gas prices is always subject to substantial short-term volatility. The rapid and very substantial reduction in natural gas prices with the advent of major new shale gas supplies was one such example of a durable shift in relative prices, and it demonstrated that power markets are well equipped to optimize smoothly in response to even rapid and major shifts in relative long-term prices faced by coal and gas plants. Under the two-rate option in the CPP, the first order effect is that both coal and natural gas power generation will face additional costs due to the need to purchase ERCs (akin to a carbon price) that are not borne by zero-emitting generators, and zero-emitting generators will receive a subsidy from their ability to sell ERCs. This will increase the relative cost of fossil generation relative to zero-emitting generation and change the relative costs between coal and gas generation. This shift in the relative operational costs for coal and gas plants under the two-rate approach, however, will be no different than prior shifts driven by changes in the relative price of coal and gas fuels either in response to market forces or through regulation such as in the carbon markets in California and New England, and power markets and generators will similarly respond by simply re-optimizing their operations. Similarly, the subsidy for zero-emitting generating is no different in principle than existing subsidies such as the Production Tax Credit or various state Renewable Portfolio Standards which are already well-incorporated into power markets.

From: Victor, Meg
Sent: Thursday, July 30, 2015 1:21 PM
To: Evans, DavidA; Marten, Alex
Subject: FW: Summary of Interagency Comments Under EO 12866 and 13563 -- GHG Mitigation TSD BB1

Hopefully you've already seen these? Ex 5
Ex 5

From: Harvey, Reid
Sent: Thursday, July 30, 2015 10:55 AM
To: Adamantiades, Mikhail; Victor, Meg
Subject: Fwd: Summary of Interagency Comments Under EO 12866 and 13563 -- GHG Mitigation TSD BB1

Begin forwarded message:

From: "Szabo, Aaron" <Aaron_L_Szabo@omb.eop.gov>
To: "Vasu, Amy" <Vasu.Amy@epa.gov>, "Culligan, Kevin" <Culligan.Kevin@epa.gov>, "Harvey, Reid" <Harvey.Reid@epa.gov>
Cc: "Frey, Nathan J." <Nathan_J._Frey@omb.eop.gov>, "Grossman, Andrea" <Andrea_L_Grossman@omb.eop.gov>
Subject: RE: Summary of Interagency Comments Under EO 12866 and 13563 -- GHG Mitigation TSD BB1

Attached please find a summary of interagency comments under EO 12866 and 13563 on the following documents:

1. Summary of Interagency Comments on the CPP RIA

2. Summary of Interagency Comments on the FP preamble; to note, the interagency comments are based on a previous version (the version before the one sent last night). Trying to merge the versions made distinguishing the interagency comments more difficult and was thus not provided.

Thank you and please let know if you have any questions.

Aaron L. Szabo

Policy Analyst

Office of Management and Budget

202-395-3621

Aaron_L_Szabo@omb.eop.gov